Low-Power Circuits for Brain–Machine Interfaces

Rahul Sarpeshkar, *Senior Member, IEEE*, Woradorn Wattanapanitch, *Student Member, IEEE*, Scott K. Arfin, *Student Member, IEEE*, Benjamin I. Rapoport, *Student Member, IEEE*, Soumyajit Mandal, *Student Member, IEEE*, Michael W. Baker, *Student Member, IEEE*, Michale S. Fee, Sam Musallam, and Richard A. Andersen

***Abstract*—This paper presents work on ultra-low-power circuits for brain–machine interfaces with applications for paralysis prosthetics, stroke, Parkinson's disease, epilepsy, prosthetics for the blind, and experimental neuroscience systems. The circuits include a micropower neural amplifier with adaptive power biasing for use in multi-electrode arrays; an analog linear decoding and learning architecture for data compression; low-power radio-frequency (RF) impedance-modulation circuits for data telemetry that minimize power consumption of implanted systems in the body; a wireless link for efficient power transfer; mixed-signal system integration for efficiency, robustness, and programmability; and circuits for wireless stimulation of neurons with power-conserving sleep modes and awake modes. Experimental results from chips that have stimulated and recorded from neurons in the zebra finch brain and results from RF power-link, RF data-link, electrode-recording and electrode-stimulating systems are presented. Simulations of analog learning circuits that have successfully decoded prerecorded neural signals from a monkey brain are also presented.**

***Index Terms*—Brain–machine interfaces, low-power, prosthetics, wireless neuroscience.**

I. Introduction

LARGE-SCALE chronic multi-electrode neural recording and stimulating systems have emerged as an important experimental paradigm for investigating brain function. Experiments using such brain–machine interfaces (BMIs) have shown that it is possible to predict intended limb movements by analyzing simultaneous recordings from many neurons (see [1] for a report of the first human trials of such devices, and see [2], [3] for recent reviews of this field). These findings have suggested a potential approach for treating paralysis and other disorders and disabilities in humans. Other BMIs, such as deep brain stimulators for Parkinson's disease and visual prostheses for the blind, function primarily via neural-stimulation circuitry rather than via neural-recording circuitry. BMIs for epilepsy will need both recording and stimulating circuitry, and progress toward systems with such dual capabilities has recently been demonstrated in the context of general-purpose multi-electrode arrays for experimental neuroscience [4], [5]. Chronic use of BMIs with large numbers of electrodes requires ultra-low-power operation so that the systems are miniature and implantable, heat dissipated in the brain is minimized, and frequent battery replacement and repeat surgeries associated with implanted systems are unnecessary. In this paper we describe low-power circuits that can be applied to many BMIs, focusing first on those for recording applications and then on those for stimulating applications.

Low-power neural amplifiers are extremely important in recording BMIs since one such amplifier is needed for each electrode. In this work, we first describe and present data from a micropower neural amplifier that is the most power-efficient and lowest-power differential neural amplifier reported to date, achieving an energy efficiency near the limits set by theory [6]. Although single-ended amplifiers have been shown to be capable of even better efficiency, they are significantly less effective at rejecting power-supply and common-mode noise and are thus considerably less practically suited for neural recording [7]. We then describe a novel scheme for adapting the noise floor of a neural amplifier to the noise-floor requirements at each recording site, which potentially enables multi-electrode arrays to reduce recording power by an order of magnitude.

RF data telemetry is necessary to communicate information wirelessly to and from neurons in the brain through the skull and skin. Due to the relatively high power costs of transcutaneous data communication (for example, transmitting 12-bit neural signals sampled at 20 kHz from 100 electrodes yields a data rate of 24 Mbs^{-1} and power consumption on the order of 10 mW), some form of data compression is needed to reduce the bandwidth of information transmitted from the brain. Adaptive, learnable, multi-input-multi-output neural decoding techniques that project firing-rate neuronal data onto several motor output parameters have successfully been used to decode movement intentions from neural signals (a variety of approaches are reviewed in [2], [3]). Simple linear decoding filters have proven useful for interpreting population codes of neurons in various brain regions, and these decoders perform comparably to adaptive Kalman filters and other probabilistic decoding techniques; the Appendix to [8] demonstrates that probabilistic decoders

Manuscript received November 01, 2007; revised May 16, 2008. Current version published October 24, 2008. This work was supported in part by a grant from the McGovern Institute Neurotechnology Program (MINT) at MIT. This paper was recommended by Associate Editor M. Sawan.

R. Sarpeshkar, W. Wattanapanitch, S. K. Arfin, S. Mandal, and M. W. Baker are with the Analog VLSI and Biological Systems Group, Research Laboratory of Electronics, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139 USA (e-mail: rahuls@mit.edu).

B. I. Rapoport is with the Analog VLSI and Biological Systems Group, Research Laboratory of Electronics, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (MIT), Cambridge, MA, 02139 USA, and also with Harvard Medical School, Boston, MA 02115 USA.

M. S. Fee is with the McGovern Institute for Brain Research and the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139 USA.

S. Musallam and R. A. Andersen are with the Division of Biology, California Institute of Technology, Pasadena, CA 91125 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBCAS.2008.2003198

with linear priors are equivalent to linear-filter-based decoders. In this work, we discuss how we may use a novel analog decoding and learning architecture to compute such a filter in a power-efficient analog fashion, thereby allowing high data compression (an output rate of 3×100 Hz $\times$ 8 bits $= 2.4$ kbs^{-1} is more than sufficient for achieving neural control over three motor degrees of freedom). We present circuit simulations of the architecture that successfully learn and decode a monkey's intention to move from its prerecorded neural spiking signals.

Then we describe and present experimental results from a novel low-power impedance-modulation technique for BMIs that can be used to transmit RF data from the brain flexibly and efficiently so that the power costs of communication are almost solely borne by external RF circuitry outside the skin and skull, rather than by the implanted RF circuitry within the cranial cavity as in prior designs [9]–[11]. Thus, heat and power dissipation in implanted hardware within the brain can be minimized. The link achieves 0.65 nJ bit^{-1} communication efficiency for forward and reverse telemetry at 2 cm distances, among the most efficient reported [12]. The forward uplink bandwidth from the brain enables 5.8 Mbs^{-1} data rates and the reverse downlink bandwidth to the brain enables 300 kbs^{-1} data rates. The external unit consumes at most 2.5 mW of power, while the internal unit consumes only 140 μW of power in the worst case.

Efficient wireless links that transmit RF power through the skin are necessary to power implanted chips in the interface directly via rectification and possibly also to recharge implanted batteries capable of a finite number of recharges. In this work we present data from an RF power link that achieves efficiencies near that set by theory for links in the 1–10 mm range (74% and 54% efficiencies) and summarize the tradeoffs needed to optimize such links [13].

In addition to being small and extremely power efficient, practical brain–machine interfaces also need to be programmable and capable of reporting high-bandwidth analog or lower-bandwidth spiking information from a few neurons if needed, sorting spikes from multiple neurons, and operating robustly in RF and mixed-signal environments. In this work we propose a novel low-power mixed-signal architecture for such interfaces that combines the power efficiency of an implanted programmable analog system with the flexibility of an external digital processor such that efficiency and flexibility are simultaneously achieved by combining the best of the analog and digital worlds.

Certain BMIs, such as visual prostheses for the blind, systems for treatment of epilepsy or Parkinson's disease, or experimental systems for neurophysiology, may require neural stimulation rather than (or in addition to) neural recording. We present experimental results from a chip used for wireless stimulation of neurons in a zebra finch brain and discuss how simple wake-up circuitry can be used to reduce power consumption in such systems.

This paper is organized as follows: In Section II we discuss the adaptive micropower neural amplifier and its use in multi-electrode systems. In Section III we discuss the analog linear decoding and learning architecture. In Sections IV and V we discuss RF data and power links, respectively. In Section VI we discuss a mixed-signal architecture for BMIs that can enable efficiency and flexibility. In Section VII we discuss wireless neural-stimulation circuits. In Section VIII we conclude by summarizing our contributions.

Brief descriptions of preliminary results of this work were presented at a conference from which some papers in this special-issue journal (including the present article) were selected [14]. Since that conference, details on the RF power link and the micropower neural amplifier have appeared in more specialized journal articles [13], [6]. Therefore, the description of these portions of the work will be brief in this broader paper. However, the brief descriptions will enable this paper to be self-contained and will provide context for the other work in the paper.

II. Micropower Neural Amplifier and Adaptive Power Biasing Scheme

A. Micropower Neural Amplifier

Fig. 1(a) shows the architecture of our adaptive micropower amplifier. The first gain stage is similar to that reported in [15] except that it is implemented with the use of an all-subthreshold and folded-cascode architecture shown in Fig. 1(b), allowing 2.8-V operation rather than 5-V operation. In [15], a 5-V power supply is necessary for maintaining large overdrive voltages in some above-threshold transistors to minimize their noise contributions. We add a bandwidth-limiting stage to keep the overall bandwidth constant as we vary the bias current of the gain stage to adapt its noise per unit bandwidth. The additional power of the bandwidth-limiting stage is negligible because the 100-fold gain provided by the gain stage alleviates its noise floor requirements. Fig. 1(c) shows an input-referred signal recorded *in vivo* from the RA region of the brain of an anesthetized zebra finch bird using a Carbostar 800 kΩ impedance electrode and our amplifier. The trace exhibits no discernible difference when compared with a recording made using a commercial neural amplifier.

We measured an input-referred noise of 3.06 μV rms over a -3 dB bandwidth of 45.0 Hz–5.32 kHz with a power consumption of 7.56 μW for 40.8 dB of mid-band gain. The noise contributions of our amplifier are minimized to be almost those of only its two input transistors, due to the use of cascoded resistive loading rather than current-source loads. Thus, the measured Noise Efficiency Factor (NEF) is 2.67, very near the theoretical limit of 2.02 for a differential amplifier, representing the most power efficient and lowest-power differential design to date. Further details of the amplifier are described in [6] including use of this amplifier for Local Field Potential (LFP) recording.

B. Adaptive Power Biasing of Neural Amplifiers in Multi-Electrode Arrays

The power required to build an amplifier with constant bandwidth, constant power-supply voltage, and an input-referred noise v_n scales as $1/v_n^2$ if the amplifier's minimum detectable signal is limited by thermal noise. This relation clearly shows the steep power cost of achieving low-noise performance in thermal-noise-limited amplifiers. Most neural amplifiers are carefully designed to be thermal-noise limited rather than

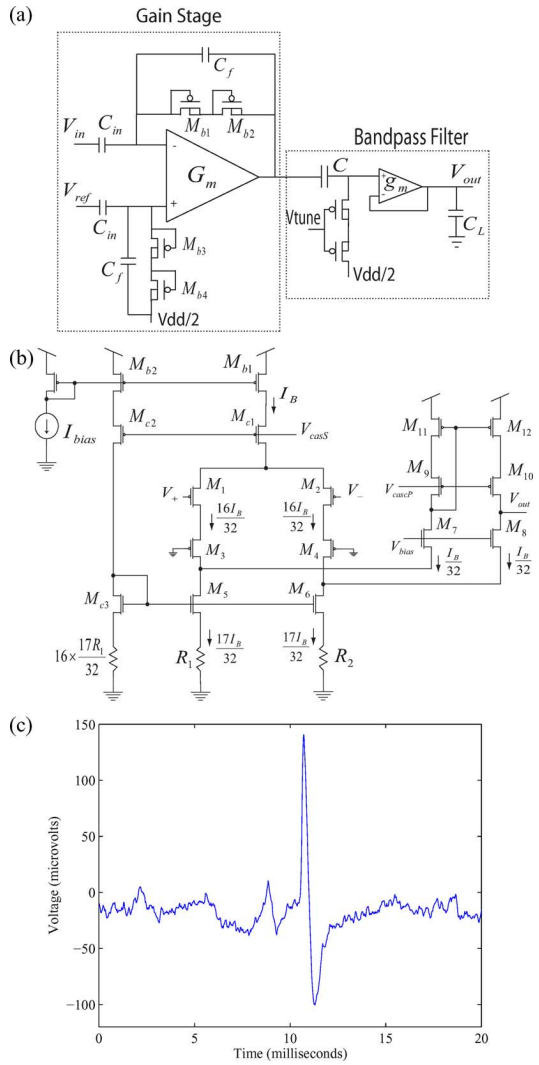

Fig. 1. (a) Overall system diagram of the neural amplifier. (Note that C_{in} is a 15-pF poly-poly integrated capacitor.) (b) Schematic of the low-power low-noise OTA used in the neural amplifier. (c) Neural recording from the brain of a zebra finch using the amplifier described.

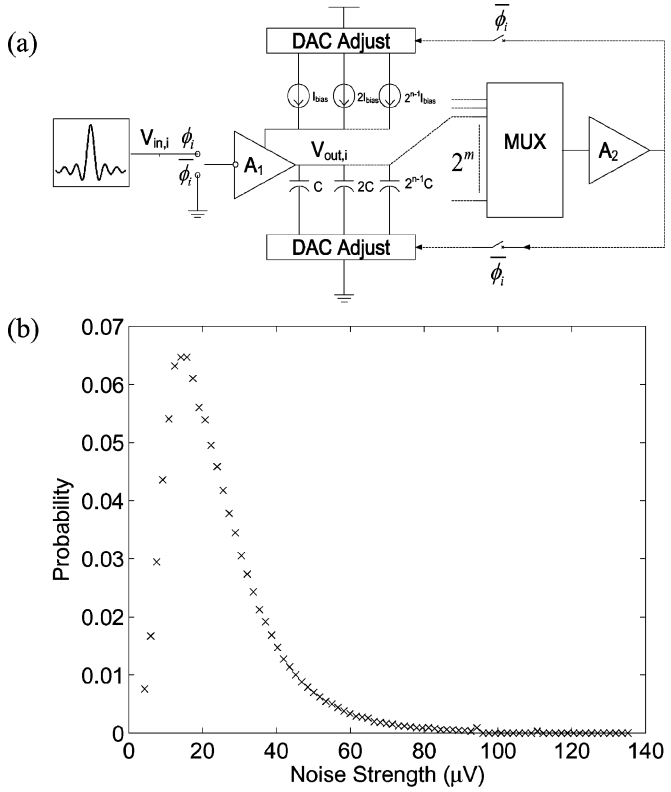

Fig. 2. (a) Schematic illustrating our adaptive biasing strategy for amplifier arrays. (b) Probability distribution of input-referred voltage noise measured from electrodes in a 64-channel array implanted in the posterior parietal cortex of a rhesus monkey for chronic neural recording.

$1/f$-noise limited to achieve the best possible performance. Neural amplifiers have been designed to handle the worst-case range of signal strengths that may be expected in any recording situation. In practice, there is considerable variance in the noise and action potential strengths of typical recordings. The steep cost of achieving low-noise performance in an amplifier suggests that rather than designing amplifier arrays with the lowest noise at all locations, significant power savings can be achieved if each amplifier can adapt its input-referred noise to the local noise floor. This adaptability enables the overall power in a multi-electrode system to be determined by a typical electrode rather than by the worst-case electrode. The control loop for setting each amplifier's bias current in a multi-electrode array may easily be implemented with little power overhead per recording site: One very-low-noise neural amplifier is used infrequently to evaluate the noise floor at each recording site in a sequential and multiplexed fashion, so that its power overhead is shared amongst all recording sites and it is only active during calibration. This scheme is shown in Fig. 2(a). Fig. 2(b) shows a typical probability distribution that we obtained from neural data recorded using a chronically-implanted 64-electrode array in a rhesus monkey. For this probability distribution, adaptive power biasing yields a 12-fold reduction in neural recording power for an entire system of 100 electrodes. *Note that the technique of adaptive power biasing may be applied to any neural amplifier, not just our amplifier.* For maximum efficacy, techniques must be used to keep the bandwidth of the amplifier invariant as its bias current is changed. In our amplifier, since the first stage determines the noise per unit bandwidth and the second stage determines the bandwidth, the total output noise is controlled simply by varying the bias current of the first stage while that of the second stage is kept constant. In other amplifiers, the output capacitance and bias current can be increased in proportion with one another to reduce the total output noise while keeping the bandwidth invariant, as shown in Fig. 2(a). In our amplifier, the value of resistances in Fig. 1(b) can be scaled with a scaling resistance R such that $I_B R$ is constant as I_B is changed. Consequently, the NEF is constant with bias current level.

Fig. 3(a) illustrates the feedback loop of a circuit used to implement one instantiation of adaptive power biasing in a neural amplifier described in [14]. This neural amplifier's noise performance is slightly inferior to the one shown in Fig. 1(b). A 'command current' that is proportional to the desired noise amplitude is determined by a prior stored measurement with a low-noise neural amplifier and a wide-dynamic-range V-to-I envelope detector described in [16]. This current is input to the adaptive

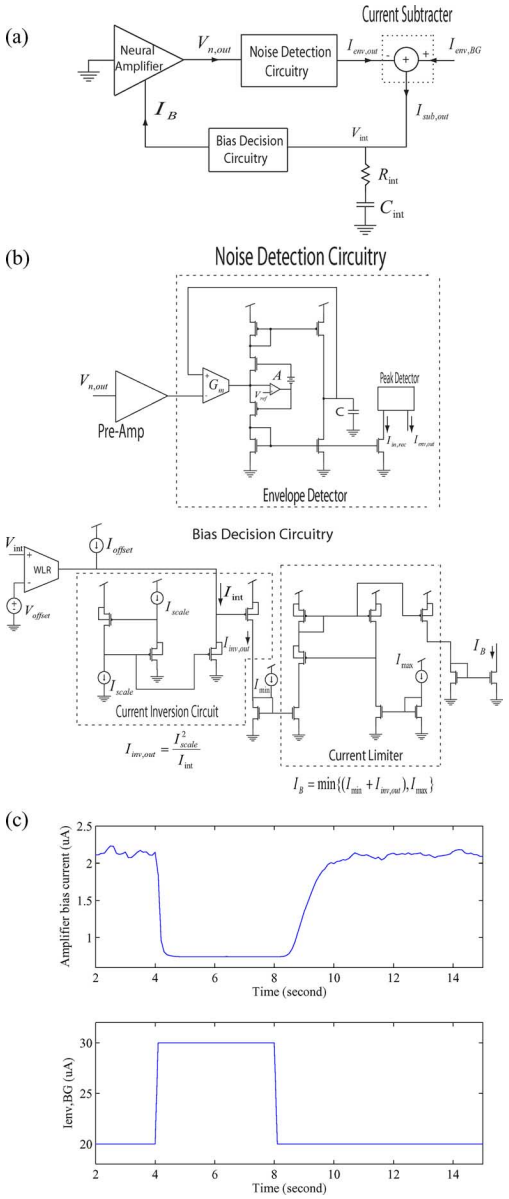

Fig. 3. (a) Block diagram of the noise calibration loop for our adaptive biasing scheme. (b) Schematics for the noise-detection circuitry and bias-decision circuitry. (c) Step response of the amplifier's bias current due to a step change in the control input current.

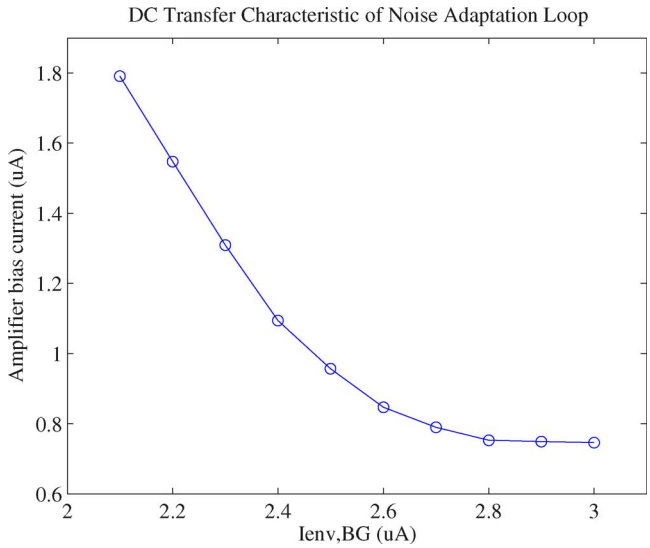

Fig. 4. Amplifier bias current of the noise calibration loop as a function of the control input current.

power-biasing circuitry. When the circuitry settles to equilibrium, the bias current of the neural amplifier being calibrated then automatically gets set to a value such that the output noise is at the command value. The use of the RC element yields a pole-zero compensation network that, together with another pole in the feedback loop, yields good phase margin in the loop. Fig. 3(b) illustrates the circuits used to implement the blocks of Fig. 3(a). Fig. 3(c) illustrates the dynamic settling performance of the adaptive-biasing loop. Fig. 4 shows that the bias current of the amplifier varies with the command input in a reciprocal fashion such that large envelopes reduce the bias current and vice versa. In this instantiation, low-leakage sample-and-hold techniques used to construct very-long-hold-time analog memories (8 bits for 9 hours) with ultra-low leakage of 5 electrons per second [17], [18] could be used to store the bias current value on a capacitor between calibrations. In other instantiations, current-DACs and digital bits could store the bias-current value as in Fig. 2(a).

Analysis of the digitized output of a neural amplifier followed by digitally controlled setting of the amplifier's bias current can implement more sophisticated adaptive biasing: For example, such schemes could set the amplifier's power at a low value if there are large action potentials on a particular electrode even though its noise floor is low. In such a case, needlessly low-noise amplifiers are not required and we can exploit this knowledge to save power. Thus, *the key idea behind adaptive power biasing is not the exact control algorithm or circuit that is used to bias the amplifier but the fact that biasing can be determined by knowledge of the statistics of the array rather than by worst-case assumptions about its statistics*. In arrays with large numbers of electrodes, the savings in power with adaptive biasing exceed an order of magnitude because outlier or low-probability points are prevented from scaling power quadratically according to worst-case needs in all amplifiers.

III. Analog Decoding and Learning Circuits for Data Compression

In some of our prior work on a bionic-ear (cochlear implant) processor, we experimentally demonstrated that analog pre-processing and delayed digitization enable order-of-magnitude power reductions over traditional A-to-D-then-DSP implementations [19], [20]. Such implementations can also preserve programmability, as well as robustness to offset, power-supply-noise, and temperature variations. In this work, we investigated whether it may be possible to achieve similar power reductions in implementing digital linear decoding and learning algorithms with analog architectures operating on analog neuronal firing rates.

Our novel analog decoding architecture uses a continuous-time, adaptive linear filtering algorithm to map neural signal inputs onto motor command outputs. The system is an m to n convolutional decoder that accepts n preprocessed neural signals

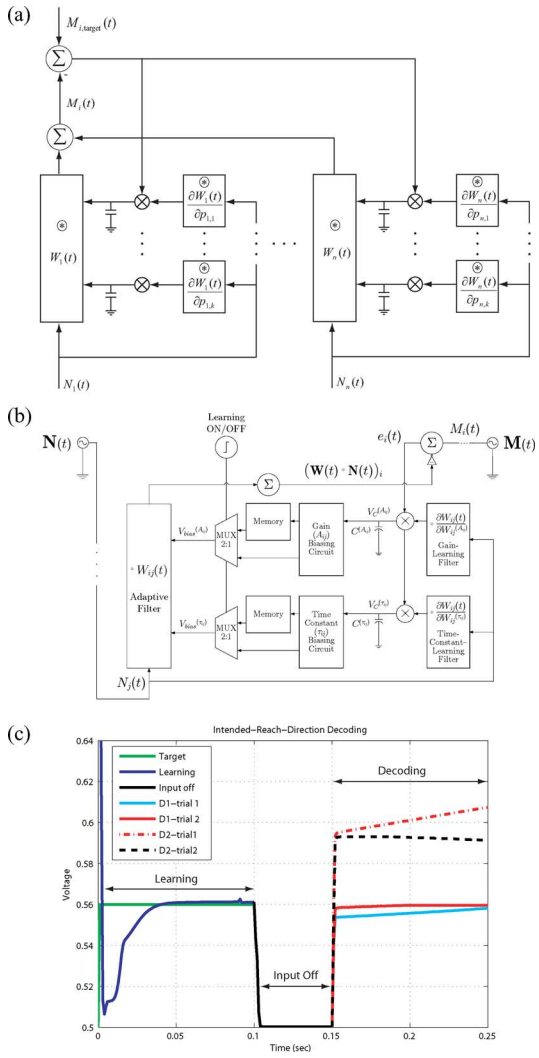

Fig. 5. (a) Block diagram illustrating the learning process used in real time to optimize the parameters of the adaptive filters that decode neural signals. (b) Block diagram of the circuit modules used to implement the neural decoding architecture. (c) Example performance of the analog learning architecture in one learning and two decoding trials. The neural input signals to the decoder, which are not shown here, were previously recorded neural signals from the posterior parietal cortex of a rhesus monkey trained to make reach movements in a stimulus-response behavioral paradigm. Optimization of the decoding kernels requires approximately twenty to thirty such learning trials, and decoding performance is evaluated on the basis of many decoding trials of the kind shown here.

$N_i(t)$ as inputs and produces m motor control signals (reach directions), $M_i(t)$ as outputs. Fig. 5(a) shows the architecture for a single motor output signal. The $m \times n$ array of adaptive filters $W_{ij}(t)$ forms a set of convolution kernels between the inputs and outputs that is analogous to the matrices of 'synaptic weights' used in artificial neural networks ($W_{ij}(t)$ is convolved with the jth neural input to form the jth component of the sum that yields the ith motor output). The parameters $p_{i,j}$ of these filters are set during a supervised learning phase by using a continuous-time analog gradient descent algorithm. Further mathematical detail lies beyond the scope of this paper, but we note here that the algorithm results in a modified version of the 'delta' learning rule well known in machine learning. Our algorithm exhibited accuracy comparable to that of the Bayesian (probabilistic) decoder described in [21] operating on the same signals (real neural recordings), but was far less computationally intensive.

Using simple continuous-time transconductor-capacitor (G_mC) filters along with multiplier, adder, and subtractor circuits in subthreshold technology as shown in Fig. 5(b), we were able to show via SPICE circuit simulations in a 0.18 μm process that our decoding architecture is capable of learning and then decoding a monkey's intention to move its arm in particular directions on the basis of data recorded from ten neurons in its posterior parietal cortex. Spike-time data from the monkey cortex recorded during short intervals just preceding reach movements were converted into analog firing rates using fourth-order wavelet-like analog matched filters on 1-ms-wide spiking inputs, followed by thresholding, followed by third-order analog interpolation filters. The architecture is able first to learn how to decode the intended direction of motion of a monkey's arm from input spike-time data and reduce the mean-squared error between the intended output (the executed reach) and the prediction made by the circuit. This is accomplished by gradient-descent-based tuning of parameters defining the convolution kernels of the adaptive decoding filter in real time. After the end of the learning period, the parameters of the decoding filter are fixed and the decoder is able to predict the intended direction of motion of the monkey's arm on the basis of new neuronal control signals. This process is demonstrated in Fig. 5(c). The analog filters were trained during the interval labeled "Learning" using the neural data observed just before the monkey reached in direction D1. Once training was completed, all the learned filter parameters were stored. After training, therefore, the filters were least-squares optimized and configured for mapping the neural signal that the monkey used for moving in the D1 direction to an output voltage of 0.5 V. We demonstrate a test of the accuracy of the learned mapping during the interval labeled "Decoding" in which the neural signals for moving in directions D1 and D2 were applied to the analog decoder. The plot shows that when the neural signal for moving in the D1 direction is applied the output voltage settles at 0.5 V, whereas it settles elsewhere when the neural signal for D2 is applied. This result indicates that the analog decoder has learned the appropriate mapping from neural signals to the intended reach direction.

The power consumption of a single channel is near 54 nW and a complete 100-channel system with 3 motor outputs is implementable with 17 μW using a 1-V power supply on a modest-sized chip. Thus, our analog architecture is extremely power efficient. It can enable a dramatic reduction in communication power due to its data-compression properties: 24 Mbs^{-1} of uncompressed data in (100 channels sampled at 20 kbs^{-1} with 12-bit resolution) to 2.4 kbs^{-1} of compressed data out (3 motor parameters at 100 bs^{-1} with 8-bit resolution). An actual chip implementation would need overhead for offset and mismatch compensation via DAC calibration, and temperature- and power-supply-immune biasing, as our prior work has shown [19], [20]. Much of the power savings results from the fact that the computational tasks can be mapped very efficiently to a slow-and-parallel analog architecture. The use of analog circuits

to perform compression saves power in data telemetry circuits due to the reduced bandwidth needs for communication. It also saves power in analog-to-digital conversion circuits that can be operated at significantly reduced bandwidths.

Our architecture represents the ultimate form of compression in the sense that the decoding computation is performed on site and only the results are transmitted. It is only viable because analog implementations of decoding computations needed in motor prosthetics can be done so efficiently. Other forms of data compression that have been proposed before, such as adaptive thresholding followed by transmission only of spike-time information [9], [22], will need to be added to an architecture like ours to preserve general-purpose flexibility in a brain–machine interface. Our architecture reveals that, for the special-purpose needs of a paralysis prosthetic, for example, extremely power-efficient analog implementations are possible. As we discuss in Section VI, the combination of general-purpose flexible architectures that are relatively inefficient with special-purpose architectures that are efficient can be used to create a system that is both flexible and efficient.

IV. RF Data Link

Fig. 6 shows a low-power half-duplex RF data-telemetry link implemented with inductively coupled external (primary) and implanted (secondary) resonators for bidirectional communication through the body. Communication from inside the body to outside the body is termed the uplink and communication from outside the body to inside the body is termed the downlink. We designed our RF link for the case of a recording BMI in which the uplink reports information from the brain and can have a bandwidth of several Mbs^{-1}, while the downlink sends information to the brain and can have a bandwidth of several hundred kbs^{-1} in the worst case. The downlink in a recording BMI typically transmits relatively low-bandwidth control, programming, and supervisory learning information.

We shall first discuss the design of the high-bandwidth uplink. The primary external unit broadcasts a 25 MHz continuous-wave carrier using an LC oscillator with positive feedback. For the high-bandwidth uplink, the secondary implanted unit modulates its impedance by open-circuiting or short-circuiting a parallel resonator, thus modulating the reflected impedance seen at the primary. The impedance seen in the primary is modulated by a factor that we call the *modulation depth* m. This quantity is given by [13], [23]

$$m \approx k^2 Q_1 Q_2 \tag{1}$$

where k is the coupling factor between the coils, set by their geometry and separation; Q_1 and Q_2 are the quality factors of the coils; and the approximation is valid if $m \ll 1$, which is usually the case. The strong dependence of m on k makes this configuration unsuitable for long-range links since k varies with the cube of the distance between the coils. For short-range links (such as those used in neural prosthetic devices, which rarely require coil separations greater than 2 cm), however, impedance modulation has the great advantage of requiring almost no power

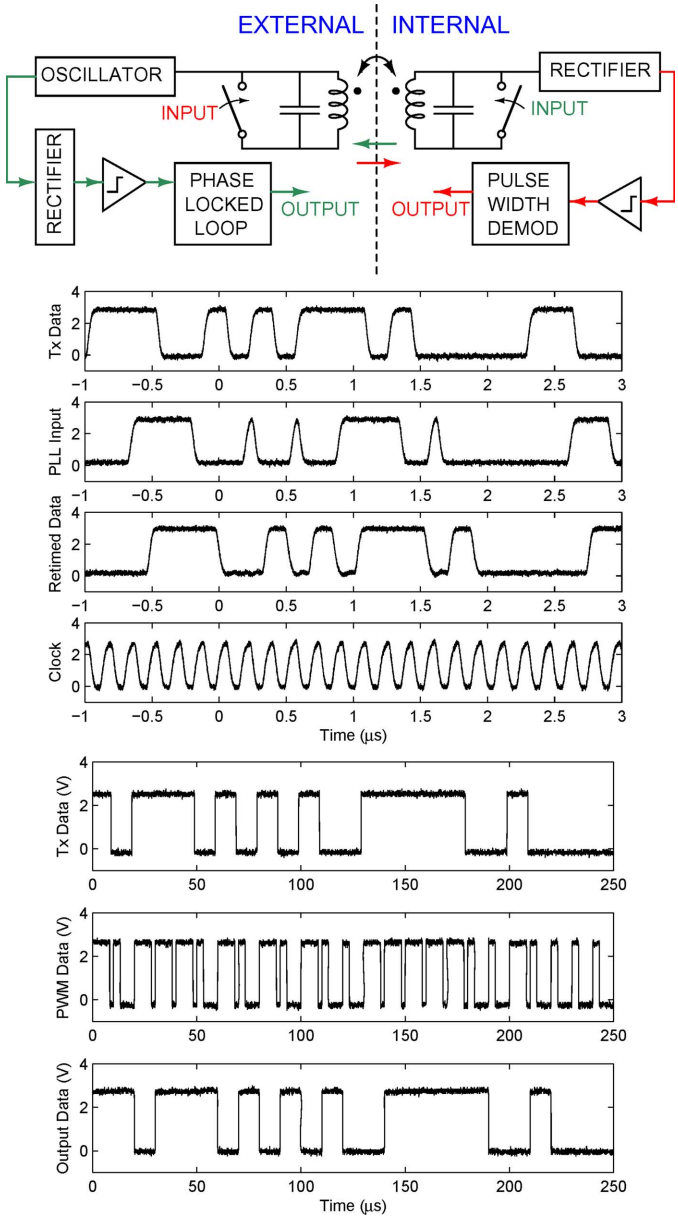

Fig. 6. (Top) Schematic of a low-power RF data-telemetry link implemented with inductively coupled external and implanted resonators for bidirectional communication through the skin. (Center) Data transmission waveforms obtained in experiments demonstrate successful recovery of uplinked data at 5.8 Mbs^{-1} with a two-bit delay between transmitter and receiver. (Bottom) Data transmission waveforms at left show successful recovery of downlinked data at 200 kbs^{-1} with a one-bit delay between transmitter and receiver.

dissipation on the secondary (internal) side, since shorting or opening a switch costs far less energy than operating a transmitter. The load modulation results in amplitude modulation of the oscillator voltage in the primary. This amplitude modulation in the primary is detected by an envelope detector built with a rectifier and lowpass filter as shown in Fig. 6. The output of the lowpass filter is thresholded by a hysteretic comparator and fed to a phase-locked loop for clock and data recovery.

For the lower-bandwidth downlink, on-off keying of the primary oscillator results in pulse-width amplitude modulation in the secondary, which is then detected by rectification, thresholding, and pulse-width demodulation circuits.

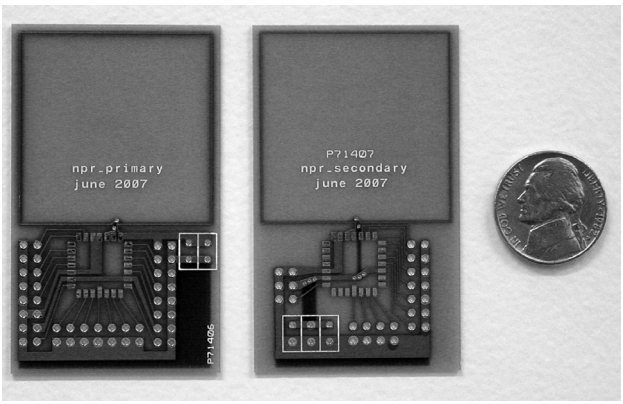

Fig. 7. Primary (external, left) and secondary (implanted, right) RF data telemetry test circuit boards.

TABLE I
PERFORMANCE SUMMARY FOR THE RF TELEMETRY DATA LINK

Performance Parameter	Value
Link Distance	tested up to $\leq$ 5 cm
Center Frequency	25 MHz
Uplink Data Rate	1 Mbs^{-1}–5.8 Mbs^{-1} (at 2 cm)
Uplink Encoding	Non-Return (NR)
Downlink Data Rate	15 kbs^{-1}–300 kbs^{-1} (at 2 cm)
Downlink Encoding	Pulse-Width Modulation (PWM)
Power supply voltages	$V_{DDH} = 2.8$ V / $V_{DDL} = 1.4$ V
External Power Consumption	2.5 mW (Uplink)/ 1.5 mW (Downlink)
Internal Power Consumption	100 μW (Uplink) / 140 μW (Downlink)
Fabrication Process	AMI 0.5 μm CMOS
Chip Size	1.5 mm $\times$ 1.5 mm (Each Transceiver)

Fig. 7 shows primary and secondary test boards that were used for making link measurements. Identical transmission and receiving coils were printed on the boards. Each coil was square, 3.5 cm on a side, and had two turns. The designed inductance was 500 nH with a simulated quality factor of 30 at 25 MHz. Packaged chips were surface mounted on the boards and aligned parallel to each other at various separations for testing. Uplink and downlink data are shown in Fig. 6 for a link separation of 2 cm. The bit error rate was $< 10^{-3}$ for the uplink at data rates $<$ 4 Mbs^{-1}, and $< 10^{-6}$ for the downlink at all tested data rates [23]. We also tested the link in the presence of a 2 cm thick layer of 0.9% saline solution between the coils. As in prior reports such as [10], the goal was to verify that the wireless link would operate normally in the presence of body tissue. No significant differences in performance were noted.

Table I summarizes the performance of the overall telemetry link including bandwidth and power consumption obtained from measurements on the link. We note that the implanted power is minimal for both the uplink and downlink (100 μW and 140 μW, respectively), while the external unit consumes 2.5 mW and 1.5 mW in the uplink and downlink modes, respectively. The uplink data rates can be as high as 5.8 Mbs^{-1} and the downlink data rates can be as high as 300 kbs^{-1}, yielding energy efficiencies of 0.65 nJ bit^{-1}. The good energy efficiency of the link results from the use of simple and energy-efficient transmitter and receiver modulation and demodulation strategies. Since the link was not found to be thermal-noise limited, further improvements in energy efficiency with future designs are possible.

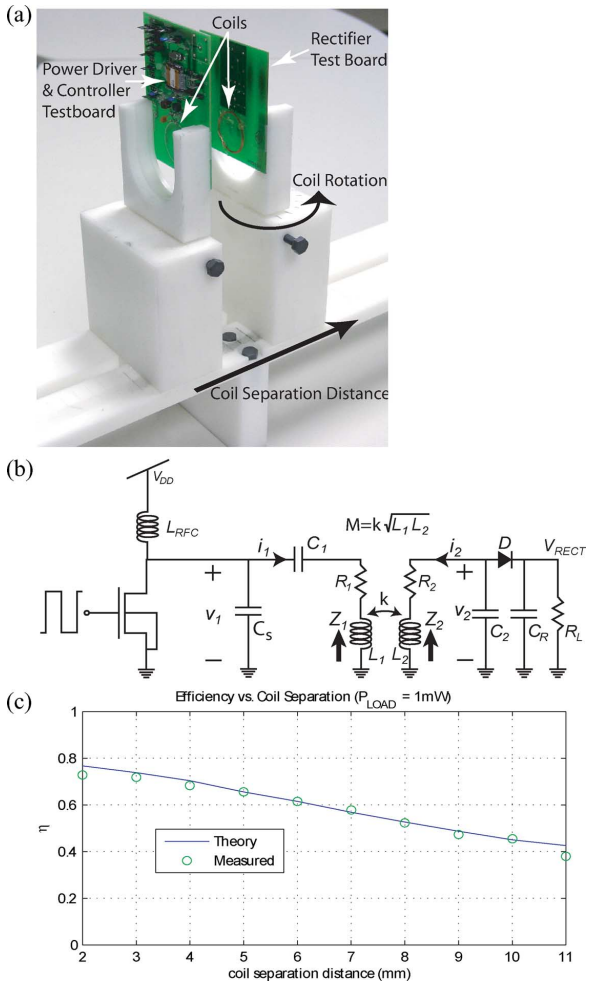

Fig. 8. (a) Testing system for RF power telemetry system, consisting of primary (external) and secondary (implanted) unit circuit boards, as well as mounts for adjusting inter-coil separations and angles. (b) Schematic of our RF power link circuits and rectifier. (c) Plot of the theoretical and measured efficiency of the RF power link as a function of distance between the primary and secondary coils.

V. RF POWER LINK

An RF power link test setup is shown in Fig. 8(a). The system uses a custom Class-E 4.5 MHz driver built on a chip, a primary resonator circuit, a secondary resonator circuit, and a Schottky diode rectifier as shown in Fig. 8(b). The rectifier loads the secondary resonator with an effective AC resistance of $R_L/2$ if there is little ripple on the load R_L, a necessary condition for a good power supply. Ignoring rectifier and driver losses, a theoretical analysis shows that the maximum possible power efficiency in the link occurs when the load R_L is chosen so that

$$Q_L = \omega \frac{R_L C_2}{2} = \frac{1}{k}\sqrt{\frac{Q_1}{Q_2}} \tag{2}$$

and is given by

$$\eta_{\max} = \frac{k^2 Q_1 Q_2}{(kQ_1+1)(kQ_2+1)} \tag{3}$$

where k is a geometry-dependent coupling factor and Q_1 and Q_2 are the quality factors of the primary and secondary resonators, respectively [13]. Fig. 8(c) shows that theoretical and experimental measurements are in good accord (the results shown here were obtained for $Q_1 \approx 40$ and $Q_2 \approx 70$). RF power links can thus be quite power efficient when operated near their optima.

In our work and in the present discussion different RF coils and RF frequencies have been used for experimental convenience, to characterize the power and data links and to explore the bounds on energy efficiency in power and data transmission systems separately. However, it must be realized that in a practical system, space constraints may require that the same coils be used for data and power telemetry, although the operating frequencies may differ for these two modes of operation. Several systems of this kind designed to transmit power and data over the same link are described in [24]–[27]. Power transfer is more efficient with high-Q coils and lower carrier frequencies where switching losses are minimized. However, high-bandwidth data transfer is easier with low-Q coils and higher carrier frequencies. Thus, operation at different carrier frequencies for power and data links is often advantageous. Hence, other work has used separate coupled resonators for the power link and data link [12], [28]. High-Q systems need periodic auto-calibration and auto-tuning feedback loops to maintain resonance.

In the low-power BMIs that we envisage in the future, the use of ultra-low-power electronics on the secondary side will in some systems enable an implanted battery with a finite number of wireless recharges, typically 1000 or fewer, to function for at least a decade: A secondary with 1-mW power consumption, for example, could function on an implanted 100-mAh battery for 30 years with 1000 wireless recharges, performing neural signal conditioning and analog decoding. In such systems data and power transmission requirements need not conflict even if only one coupled resonator is used for both the data link and the power link as power transfer only occurs during infrequent recharging. Large interfering RF voltages necessary for transmitting power to the secondary then do not corrupt sensitive low-power secondary circuits either. Implanted batteries do, however, require careful hermetic sealing, short-circuit protection, and battery-management circuits.

VI. Hybrid Analog-Digital System Design

Fig. 9 shows an overall architecture that consists of low-power implanted DAC-programmable analog circuits that are configured by a low-power external DSP or FPGA. Depending on the user's choice, the system can be configured to report raw neural data from a selected set of electrodes, single and multiunit spikes extracted via spike sorting, local field potential (LFP) data, or decoded motor parameters via a data telemetry uplink as we have previously described. Such flexibility is possible because the parameter values for the analog wavelet-like matched filters and other analog parameters are

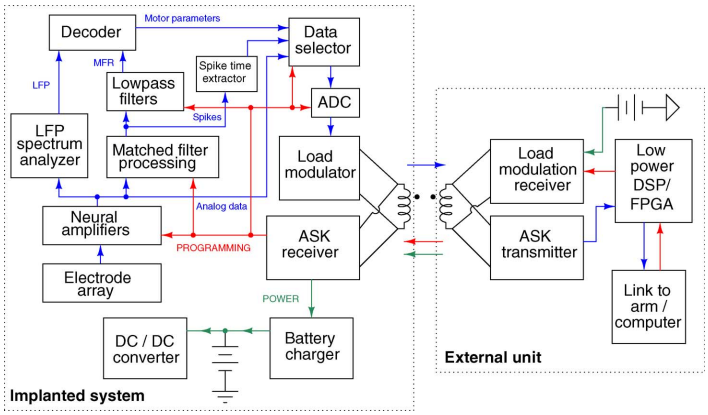

Fig. 9. Block diagram illustrating the overall architecture of a hybrid analog-digital brain–machine interface system.

determined by detailed digital analysis of raw uplinked analog neural data, then downloaded into the implant via the low-bandwidth telemetry downlink. Since the relatively power-hungry digital analysis need only be done occasionally to keep the system up-to-date as signal properties evolve (due to effects such as relative movement between neurons and electrodes), the overall power consumption of the system remains low but the flexibility of a digital system can be leveraged. Low-power implanted analog architectures for decoding that are highly efficient could have their parameters determined by flexible digital processors in an external unit and then downloaded into the implanted unit. Thus, the combination of external digital programmability combined with internal analog efficiency can enable an overall system to incorporate the best of the analog and digital worlds.

VII. Wireless Neural Stimulation

Thus far we have focused on circuits and systems capable only of recording from neurons in the brain. Fig. 10(a) shows a system that we have used for wirelessly stimulating neurons in the zebra finch brain to study the neural basis of birdsong as shown in Fig. 10(c). For example, stimulation in the zebra finch brain while the bird is singing can yield insight into the mechanisms of song generation [29]. The system consists of an external transmitter (not shown) controllable through a computer interface, and a miniature, implantable wireless receiver-and-stimulator. The miniature printed circuit board contains a receiver coil and electrodes (reverse side), battery (not shown), and a custom integrated circuit for data demodulation and neural stimulation. The chip, fabricated in a standard 0.5 μm CMOS process, occupies 2.25 mm^2 and is capable of delivering biphasic current pulses to 4 addressable electrode sites at 32 selectable current levels ranging from 10 μA to 1 mA, as shown in Fig. 10(b). The entire implant weighs 0.6 g (including the battery) and occupies a footprint smaller than 1.5 cm^2. Fig. 10(c) shows simultaneous recording from one area of the bird brain, HVC, while stimulation was applied to another area, Area X. The observed action potentials in HVC (after the stimulus artifacts due to stimulation of Area X have subsided) prove that the wireless neural stimulation system is working. Fig. 11(a) illustrates the Colpitts oscillator transmitter

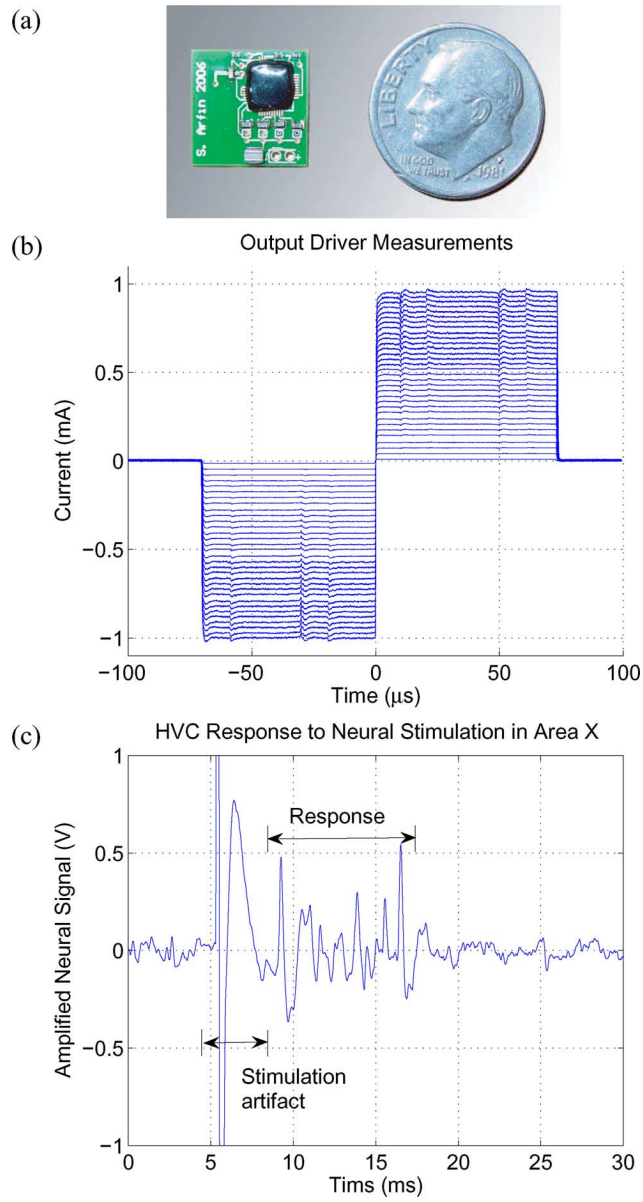

Fig. 10. (a) Photograph of the chip-on-board wireless neural stimulation system. (b) Measured biphasic current pulses demonstrating 32 programmable stimulation levels. (c) HVC response to neural stimulation in Area X of the zebra finch brain.

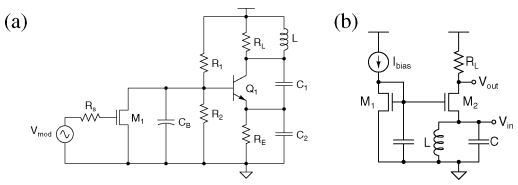

Fig. 11. (a) Colpitts oscillator circuit used in our wireless neural stimulation system. (b) Receiver circuit used in our wireless neural stimulation system.

circuit and Fig. 11(b) illustrates the receiver circuit that we used in this neural stimulation system.

A wake-up system on the chip permits us to extend the life of the battery over a full 30 days and thus facilitates extended neuroscience experiments. During periods of birdsong inactivity,

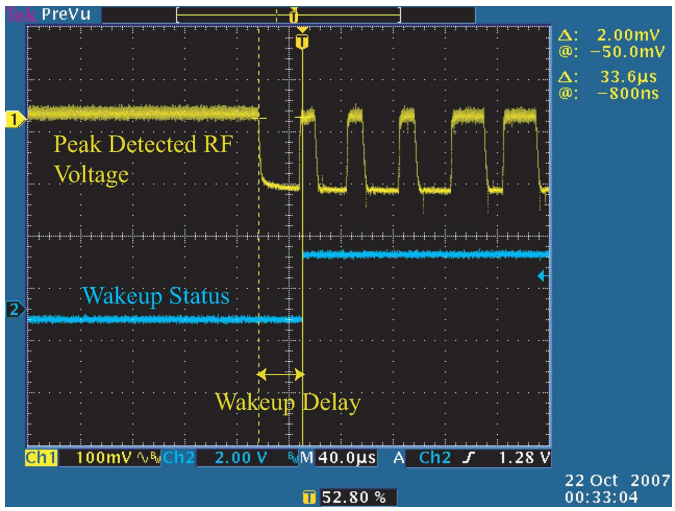

Fig. 12. Experimental waveforms illustrating the wake-up operation of the neural stimulation system.

the receiver chip enters a sleep mode in which it consumes only 7 μA of quiescent current. When an external microphone detects song, a computer activates the wireless transmitter, which generates an RF signal. A detector on the chip receives the RF signal and 'wakes up' the data demodulation and output-driver circuitry: A rectifier on the chip converts the amplitude-modulated RF signal to a full-scale digital signal, whose edge is detected via XOR circuits and used to wake up other circuits. The rectifier's time constant is determined by the system data rate, typically 25 kbs^{-1}, such that the system is capable of waking up in just a single bit period (about 40 μs). Fig. 12 shows experimental waveforms illustrating the wake-up operation of the system. After the system is assumed to have awakened, a complete stimulation command is issued. The entire time from when the bird first begins to vocalize until the device can deliver stimulation is only about 1 ms. This delay is insignificant compared to the duration of a typical songbird vocalization [30]. When the bird is not singing, the RF signal is turned off and an internal timer puts the chip to sleep in about 1 second. When the chip is awake, static power consumption in the core is about 16 μA. However, the power consumption in the output stage can be significantly greater, reaching as much as 100 μA for the reference current when full-scale stimulation currents of 1 mA are required. The use of the wake-up system thus significantly improves battery lifetime. Such feed-forward signal-triggered stimulation could be used to save power in other brain–machine interfaces; the stimulation circuits of devices used to treat epilepsy, for example, could be activated only when triggered by seizure-detection signals. Of course, such power savings are only practical if detection is much cheaper than stimulation, as is often the case. Algorithmic strategies [31], [32], switched-capacitor strategies [33], and better electrode design can also lower stimulation power. Techniques to create highly accurate charge-balancing circuits that obviate the need for large dc blocking capacitors and that consequently lower implanted-system size have also been described [34].

VIII. Conclusion

We have presented several circuits and architectures for low-power recording, processing, stimulation, and wireless transmission of neural signals in brain–machine interfaces. These include a state-of-the-art micropower differential neural amplifier; adaptive power biasing of amplifier arrays in multi-electrode systems; analog architectures for neural signal decoding, learning and data compression; 0.65 nJ bit^{-1} impedance-modulation-based bidirectional wireless links that minimize implanted-unit power consumption; RF links that achieve theoretically optimal power-transfer efficiencies; hybrid analog-digital architectures that combine flexibility and efficiency; and wireless neural stimulation circuits that exploit sleep modes to save power while allowing quick wake-up. Together, such circuits and systems could enable highly power-efficient brain–machine interfaces to be developed, thus bringing them a step closer to universal accessibility in experimental neuroscience and widespread clinical use.

References

[1] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature*, vol. 442, pp. 164–171, Jul. 2006.

[2] M. A. Lebedev and M. A. L. Nicolelis, "Brain–machine interfaces: Past, present and future," *Trends Neurosci.*, vol. 29, no. 9, pp. 536–546, 2006.

[3] A. B. Schwartz, D. J. Weber, X. T. Cui, and D. W. Moran, "Brain–controlled interfaces: Movement restoration with neural prosthetics," *Neuron*, vol. 52, pp. 205–220, 2006.

[4] E. A. Brown, J. D. Ross, R. A. Blum, Y. Nam, B. C. Wheeler, and S. P. DeWeerth, "Stimulus-artifact elimination in a multi-electrode system," *IEEE Trans. Biomed. Circuits Syst.*, vol. 2, no. 1, pp. 10–21, Mar. 2008.

[5] R. A. Blum, J. D. Ross, E. A. Brown, and S. P. DeWeerth, "An integrated system for simultaneous, multichannel neuronal stimulation and recording," *IEEE Trans. Circuits Syst. I: Fundamental Theory Appl.*, vol. 54, no. 12, pp. 2608–2618, Dec. 2007.

[6] W. Wattanapanitch, M. S. Fee, and R. Sarpeshkar, "An energy-efficient micropower neural recording amplifier," *IEEE Trans. Biomed. Circuits Syst.*, vol. 1, no. 2, pp. 136–147, Jun. 2007.

[7] J. Holleman and B. Otis, "A sub-microwatt low-noise amplifier for neural recording," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2007, pp. 45–48.

[8] W. Wu, Y. Gao, E. Bienenstock, J. P. Donoghue, and M. J. Black, "Decoding of motor cortical activity using a Kalman filter," *Neural Computation*, vol. 18, pp. 80–118, 2006.

[9] R. R. Harrison, P. T. Watkins, R. J. Kier, R. O. Lovejoy, D. J. Black, B. Greger, and F. Solzbacher, "A low-power integrated circuit for a wireless 100-electrode neural recording system," *IEEE J. Solid-State Circuits*, vol. 42, no. 1, pp. 123–133, Jan. 2007.

[10] P. Mohseni, K. Najafi, S. J. Eliades, and X. Wang, "Wireless multichannel biopotential recording using an integrated FM telemetry circuit," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 13, no. 9, pp. 263–271, Sep. 2005.

[11] M. Ghovanloo and K. Najafi, "A wide-band frequency-shift keying wireless link for inductively powered biomedical implants," *IEEE Trans. Circuits Syst. I: Fundamental Theory Appl.*, vol. 51, no. 12, pp. 2374–2383, Dec. 2004.

[12] M. Ghovanloo and S. Atluri, "A wide-band power-efficient inductive wireless link for implantable microelectronic devices using multiple carriers," *IEEE Trans. Circuits Syst. I: Fundamental Theory Appl.*, vol. 54, no. 10, pp. 2211–2221, Oct. 2007.

[13] M. W. Baker and R. Sarpeshkar, "Feedback analysis and design of RF power links for low-power bionic systems," *IEEE Trans. Biomed. Circuits Syst., Inivted Paper*, vol. 1, no. 1, pp. 28–38, Mar. 2007.

[14] R. Sarpeshkar, W. Wattanapanitch, B. I. Rapoport, S. K. Arfin, M. W. Baker, S. Mandal, M. S. Fee, S. Musallam, and R. A. Andersen, "Low-power circuits for brain–machine interfaces," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2007, pp. 2068–2071.

[15] R. R. Harrison and C. Charles, "A low-power low-noise CMOS amplifier for neural recording applications," *IEEE J. Solid-State Circuits*, vol. 38, pp. 958–965, Jun. 2003.

[16] S. Zhak, M. W. Baker, and R. Sarpeshkar, "A low power wide dynamic range envelope detector," *IEEE J. Solid-State Circuits*, vol. 38, no. 10, pp. 1750–1753, Oct. 2003.

[17] M. O'Halloran and R. Sarpeshkar, "A 10 nW 12-bit accurate analog storage cell with 10 aA leakage," *IEEE J. Solid-State Circuits*, vol. 39, no. 11, pp. 1985–1996, Nov. 2004.

[18] M. O'Halloran and R. Sarpeshkar, "An analog storage cell with 5 electron/sec leakage," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2006, pp. 557–560.

[19] R. Sarpeshkar, C. Salthouse, J.-J. Sit, M. W. Baker, S. M. Zhak, T. K.-T. Lu, L. Turicchia, and S. Balster, "An ultra-low-power programmable analog bionic ear processor," *IEEE Trans. Biomed. Eng.*, vol. 52, pp. 711–727, Apr. 2005.

[20] R. Sarpeshkar, M. Baker, C. Salthouse, J.-J. Sit, L. Turicchia, and S. Zhak, "An analog bionic ear processor with zero-crossing detection," in *Proc. IEEE Int. Solid-State Circuits Conf.*, Feb. 2005, pp. 78–79.

[21] S. Musallam, B. D. Corneil, B. Greger, H. Scherberger, and R. A. Andersen, "Cognitive control signals for neural prosthetics," *Science*, vol. 305, pp. 258–262, 2004.

[22] R. H. Olsson and K. D. Wise, "A three-dimensional neural recording microsystem with implantable data compression circuitry," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2796–2804, Dec. 2005.

[23] S. Mandal and R. Sarpeshkar, "A bidirectional wireless link for neural prostheses that minimizes implanted power consumption," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Montreal, QC, Canada, 2007, pp. 3930–3933.

[24] Y. Hu and M. Sawan, "A fully integrated low-power BPSK demodulator for implantable medical devices," *IEEE Trans. Circuits Systems I: Fund. Theory Appl.*, vol. 52, no. 12, pp. 2252–2562, Dec. 2005.

[25] P. R. Troyk and M. A. K. Schwan, "Closed-loop class-E transcutaneous power and data link for microimplants," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 6, pp. 589–599, Jun. 1992.

[26] O. Omeni and C. Toumazou, "A CMOS micro-power wideband data/power transfer system for biomedical implants," in *Proc. IEEE 2003 Int. Symp. Circuits Syst.*, May 2003, vol. 5, pp. V61–V64.

[27] G. Wang, W. Liu, M. Sivaprakasam, and G. A. Kendir, "Design and analysis of an adaptive transcutaneous power telemetry for biomedical implants," *IEEE Trans. Circuits Systems I: Fund. Theory Appl.*, vol. 52, no. 10, pp. 2109–2117, Oct. 2005.

[28] L. Theogarajan, J. Wyatt, J. Rizzo, B. Drohan, M. Markova, S. Kelly, G. Swider, M. Raj, D. Shire, M. Gingerich, J. Lowenstein, and B. Yomtov, "Minimally invasive retinal prosthesis," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2006, pp. 99–108.

[29] E. T. Vu, M. E. Mazurek, and Y.-C. Kuo, "Identification of a forebrain motor programming network for the learned song of zebra finches," *J. Neurosci.*, vol. 14, no. 11, pp. 6924–2934, Nov. 1994.

[30] A. Leonardo and M. S. Fee, "Ensemble coding of vocal control in birdsong," *J. Neurosci.*, vol. 25, no. 3, pp. 652–661, Jan. 2005.

[31] J.-J. Sit, A. M. Simonson, A. J. Oxenham, M. A. Faltys, and R. Sarpeshkar, "A low-power asynchronous interleaved sampling algorithm for cochlear implants that encodes envelope and phase information," *IEEE Transactions on Biomedical Engineering*, vol. 54, pp. 138–149, Jan. 2007.

[32] J.-J. Sit and R. Sarpeshkar, "An asynchronous cochlear-implant processor that can encode music and lower stimulation power," *IEEE Pervasive Comput.*, vol. 7, no. 1, pp. 40–48, Jan. 2008.

[33] S. K. Kelly and J. Wyatt, "A power-efficient voltage-based neural tissue stimulator with energy recovery," in *Proc. IEEE Int. Solid-State Circuits Conf.*, Feb. 2004, pp. 228–524.

[34] J.-J. Sit and R. Sarpeshkar, "A low-power blocking-capacitor-free charge-balanced electrode-stimulator chip with less than 6 nA D.C. error for 1 mA full-scale stimulation," *IEEE Trans. Biomed. Circuits Syst.*, vol. 1, no. 3, pp. 172–183, Sep. 2007.

Rahul Sarpeshkar (M'97) received the B.S. degrees in electrical engineering and physics from the Massachusetts Institute of Technology (MIT), Cambridge, in 1990 and the Ph.D. degree from the California Institute of Technology, Pasadena, in 1997.

He was with Bell Labs as a Member of the Technical Staff in 1997. Since 1999, he has been on the faculty of the Electrical Engineering and Computer Science Department, MIT, where he heads a research group on Analog VLSI and Biological Systems and is currently an Associate Professor. He holds over 20 patents and has authored more than 70 publications, including one that was featured on the cover of *Nature*. His research interests include analog and mixed-signal VLSI, biomedical systems, ultra-low-power circuits and systems, biologically inspired circuits and systems, molecular biology, neuroscience, and control theory.

Dr. Sarpeshkar has received the Packard Fellow Award given to outstanding young faculty, the Office of Naval Research Young Investigator Award, the National Science Foundation Career Award, and the Indus Technovator Award. He has also received the Junior Bose Award and the Ruth and Joel Spira Award, both for excellence in teaching at MIT. He is currently an Associate Editor of IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS.

Woradorn Wattanapanitch (S'07) received the B.S. degree (*summa cum laude*) in electrical and computer engineering from Cornell University, Ithaca, NY, in 2005, and the M.S. degree in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 2007. He is currently working toward the Ph.D. degree at MIT in the Analog VLSI and Biological Systems Group at the MIT Research Laboratory of Electronics. His research interests include low-power analog and mixed-signal VLSI for biomedical applications, and control theory.

Mr. Wattanapanitch was awarded a Royal Thai Government Fellowship in 2000.

Scott K. Arfin (S'00) received the B.S. degree (*summa cum laude*) from Columbia University, New York, in 2004 and the M.S. degree from the Massachusetts Institute of Technology (MIT), Cambridge, in 2006, both in electrical engineering. He is currently pursuing the Ph.D. degree in the Analog VLSI and Biological Systems Group at the MIT Research Laboratory of Electronics, in the field of low-power electronics for biomedical applications.

Benjamin I. Rapoport (S'07) received the B.A. degree in physics and mathematics and the M.A. degree in physics from Harvard University, Cambridge, MA, in 2003, the M.Sc. degree in mathematics from Oxford University, Oxford, U.K., in 2004, and the M.S. degree in physics from the Massachusetts Institute of Technology (MIT), Cambridge, in 2007.

He is currently a student in the M.D.-Ph.D. Program at Harvard Medical School and is jointly pursuing the M.D. at Harvard Medical School and the Ph.D. in electrical engineering at MIT. His research and professional interests include bioimplantable electronic interfaces with the brain and nervous system, biological and computational neuroscience, and clinical neurosurgery.

Soumyajit Mandal (S'01) received the B.Tech. degree from the Indian Institute of Technology, Kharagpur, India, in 2002, and the M.S. degree in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 2004, where he is currently working toward the Ph.D. degree. His research interests include nonlinear dynamics, low-power analog and RF circuit design, antennas and biochemical networks.

Mr. Mandal was awarded the President of India Gold Medal in 2002.

Michael W. Baker (S'04) received the B.S., M.Eng., and Ph.D. degrees in electrical engineering and computer science in 2000, 2002, and 2007, respectively, all from the Massachusetts Institute of Technology (MIT), Cambridge.

His research interests include neural and bionic implants, low-power integrated analog design, and integrated radio-frequency circuits.

Michale S. Fee received the B.S. degree in engineering physics from the University of Michigan, Ann Arbor, and the Ph.D. in applied physics from Stanford University, Palo Alto, CA.

He joined Bell Laboratories as a Postdoctoral Fellow in the Biological Computation Research Department, and continued there as a Member of Technical Staff until 2003, when he joined the faculty at the Massachusetts Institute of Technology (MIT), Cambridge. He is presently an Investigator at the McGovern Institute for Brain Research as well as an Associate Professor in the Department of Brain and Cognitive Sciences, both at MIT. His laboratory focuses on the biophysical and neural circuit mechanisms underlying the learning and generation of sequences in the brain, with a particular interest in vocal sequences of the songbird. The Fee Laboratory also develops new technologies for measuring and manipulating neural activity in behaving animals.

Sam Musallam is an Assistant Professor in the Department of Electrical and Computer Engineering and an Associate Member of the Department of Physiology, both at McGill University, Montreal, QC, Canada.

He leads the Neural Prosthetics Laboratory at McGill University, which investigates the neuro-scientific aspects of developing optimal neural prosthetic devices, and also develops implantable devices for measurement of biological signals in the brain and body.

Richard A. Andersen is the James G. Boswell Professor of Neuroscience at the California Institute of Technology (Caltech), Pasadena.

Since 1994, he has also been the Director of the Sloan Center for Theoretical Neurobiology at Caltech. He studies the neurobiological underpinnings to such brain processes as sight, hearing, and action planning. His laboratory is also studying neural prosthetics, electronic interfaces with the brain that may eventually enable disabled humans to achieve thought-based control of external devices.

Professor Andersen is a Fellow of the American Association for the Advancement of Science and a Member of the American Academy of Arts and Sciences.